Volume 13 Number 1 1985 Nucleic Acids Research

Structural features and restricted expression of a human α-tubulin gene

John L.Hall and Nicholas J.Cowan

Department of Biochemistry, New York University School of Medicine, 550 First Avenue, New York, NY 10016, USA

Received 17 October 1984; Revised and Accepted 30 November 1984

ABSTRACT

The nucleotide sequence of a human α-tubulin gene (bα1) is described. This gene is extensively homologous to a rat α-tubulin gene in its coding regions, 3'-untranslated region and, indeed, in segments of its largest intron. However, with the exception of three short conserved blocks of homology, the 5' flanking regions of the rat and human genes are unrelated. Hence, these genes each encoding an identical protein are transcribed under the influence of divergent promoters.

Blot analyses using RNA from a variety of transformed cells derived from different tissues indicate that expression of the human α-tubulin gene is restricted to cells of neurological origin. Among neurological cell types bα1 expression is further restricted to adherent cells that are morphologically differentiated. The data presented suggest that the bα1 gene encodes a prominent neuronal and glial α-tubulin and that bα1 expression is a function of the differentiated state of these cells.

INTRODUCTION

Microtubules are filamentous structures that are present in virtually all eukaryotic cells. They are principal components of the cytoskeleton, of the mitotic spindle, of cilia and flagella, and of neuronal processes. They are involved in numerous functions including maintenance of cell shape, mitosis, cell movement, and intracellular transport (1). The major components of microtubules are the α- and β-tubulins. Heterodimers of these proteins form subunits that are polymerized in microtubules.

In mammals, the genes that encode α- and β-tubulin proteins are members of two distinct multigene families (2). These families each contain 15-20 members and include pseudogenes as well as functionally expressed sequences. Analysis of the human β-tubulin gene family has revealed that pseudogenes, most of which are of the processed type, account for the majority of sequences in the family (3,4). The high ratio of pseudogenes to functional genes that is characteristic of the β-tubulin gene family is likely to apply to the human α-tubulin gene family as well. The ongoing task of identifying

all of the functional human tubulin genes has resulted in the complete characterization of three β-tubulin genes (4-6), and, with this report, the characterization of a single α-tubulin gene.

Because tubulins are ubiquitous in eukaryotic cells and because of their role in diverse cellular functions, an important question concerns the pattern of tubulin gene regulation. There is evidence for multiple controls of tubulin gene expression. Experiments using drugs that depolymerize microtubules have indicated that the level of unpolymerized tubulin itself regulates the level of tubulin mRNA (7,8). Studies of tubulin gene expression in the rat and chicken employing 3'-untranslated region gene-specific probes have revealed that tubulin genes are differentially regulated during development (9,10). Of particular interest are a variety of experiments that demonstrate regulation of tubulin gene expression at the level of tissue specificity. These experiments, based on RNA blot transfer analysis as well as on genetic analysis of Drosophila melanogaster, indicate the following: 1) testis-specific expression of a Drosophila β-tubulin gene (11,12); 2) dominant expression in the testis of a mouse α-tubulin gene (13) and a chicken β-tubulin gene (14); and 3) occurrence of two neural specific β-tubulin mRNAs in the rat (9). In humans, RNA blot analyses using 3'-untranslated region gene-specific probes have revealed co-expression of at least two β-tubulin genes in many transformed cell lines and the restricted expression of a third β-tubulin gene to cells of neurological origin (6). Here we report the isolation and characterization of a human α-tubulin gene that also appears restricted in its expression to cells of neurological origin. Comparison of this gene with a closely related rat α-tubulin gene (15) reveals extensive homologies not only within the coding region, but within untranslated regions and intervening sequences as well. By contrast, the promoter regions of these two genes are highly divergent. The evolutionary implications of these observations are discussed.

MATERIALS AND METHODS

Library Screening, Cloning, and DNA Sequencing

The 3'-untranslated region probe of the bα1 cDNA was subcloned as described (16). This probe, labeled with ^{32}P by nick-translation was used to screen a partial Hae III/Alu I recombinant human genomic library (17). The regions within the isolated phage containing α-tubulin sequences were identified by restriction digestion and Southern blotting using an α-tubulin coding region probe derived from the kα1 cDNA (16). α-tubulin hybridizing

regions, consisting of three Eco RI fragments, were subcloned into the pKY2700 vector, a plasmid that can be positively selected for Eco RI inserts when it is used to transform NS872 cells (18).

DNA sequencing was performed by the dideoxy chain terminator method of Sanger et al. (19) using fragments subcloned into bacteriophage M13 mp8 or mp9 as templates (20). Fragments for M13 subcloning were generated by either of two methods: 1) direct restriction digestion at appropriate sites for insertion at the M13 polylinker; 2) treatment of a linear fragment with exonuclease Bal 31 to shorten the fragment, followed by a second restriction digestion and ligation into M13. The times of Bal 31 digestion were varied from 2 to 15 minutes to generate overlapping clones. A set of clones was selected on the basis of their dideoxy A tracks such that their sequences would encompass the entire fragment.

The computer programs of Staden were used for compilation of the DNA sequences (21). The DNA sequence analysis programs of Pustell were used for the matrix homology plots (22).

Cell Cultures, RNA Preparation, and Blot Analyses

Human retinoblastoma (Y79) and neuroblastoma (IMR 6, CHP 126, CHP 134d) cells were provided by Dr. Fred Gilbert of the Mount Sinai School of Medicine. The SKN-SH and IMR 32 neuroblastoma cell lines were obtained from Dr. June Beidler of Sloan-Kettering Institute. Cells were cultured in RPMI 1640 containing 10°/∘ fetal calf serum, 150 units/ml penicillin, and 100 µg/ml streptomycin.

RNA was prepared according to the procedure of Cleveland et al. (2). For RNA blot transfer experiments, 5 µg samples of polyA$^+$ RNA were run on denaturing agarose gels containing 2.2M formaldehyde (23) and the gel contents transferred to nitrocellulose by the method of Southern (24). Blots were prehybridized for 4 hrs in 10 x Denhardt's solution, 5 x SSC, 20 mM PO_4, pH 6.5 (1 x SSC = .15M NaCl, .015M sodium citrate) at 42° and hybridized for 14 hrs at the same temperature in 50°/∘ formamide, 1 x Denhardt's solution, 5 x SSC, 20 mM PO_4. Probes were ^{32}P-labeled by nick-translation (25). Blots were washed to a final stringency of 0.2 x SSC, 0.1°/∘ SDS at 68°.

S_1 Protection

A single-stranded 5'-end labeled fragment for S_1 protection was prepared as follows. A 522 bp Eco RI-Sma restriction fragment was purified by polyacrylamide gel electrophoresis and recovered from the gel by electroelution. This fragment was incubated with Exo III under conditions

specified by the manufacturer (Bethesda Research Laboratories, Inc.). The fragment was ^{32}P-labeled with polynucleotide kinase and 3 x 10^5 cpm were incubated for 16 hrs in a reaction containing 5 µg of IMR 6 polyA$^+$ RNA, 0.3M NaCl, 0.2 mM EDTA, 0.1M Tris-HCl (pH 7.5). After annealing, the reaction was diluted into 0.4 ml of 0.3M NaCl, 0.1M sodium acetate (pH 4.5), 3 mM $ZnCl_2$ and digested with 2000 or 3000 units of S_1 nuclease. The products of this digestion were analyzed on a 5°/• polyacrylamide sequencing gel.

RESULTS

Isolation and Structural Features of the Human α-Tubulin Gene bα1.

A human α-tubulin cDNA clone (bα1) that was isolated from a human fetal brain cDNA library has been previously described (16). The subcloned 3'-untranslated region of this cDNA, which detects a subfamily containing only two of the 15-20 α-tubulin sequences present in the human genome, was used to screen a human genomic library cloned in bacteriophage lambda. Two types of positively hybridizing clones were obtained and analyzed. One set of clones contained a 1.2 kb region that hybridized with an α-tubulin coding region probe. This segment was subcloned and sequenced. Comparison of this sequence with the sequence of the bα1 cDNA revealed the presence of numerous base substitutions and at least one insertion. This insertion constitutes a frame-shift mutation and results in the occurrence of several termination codons in the new reading frame. These termination codons, which would interfere with the translation of a functional α-tubulin polypeptide, are characteristic of pseudogenes. The second set of overlapping clones contained a more extensive region of α-tubulin hybridizing sequences which were identified in a 4.1 kb segment that was subcloned and sequenced according to the strategem diagrammed in Fig. 1. Comparison of this sequence with the bα1 cDNA sequence revealed complete homology, and enabled a complete definition of the structure of the bα1 gene.

The sequence data show (Fig. 2) that the gene consists of four exons interrupted by three introns. The largest intron which is 1.5 kb in length occurs between the first and second exons. This intron is located between the first and second codons of the gene and so demarcates a protein domain consisting of the initiator methionine residue alone. The second intron is located between the first and second nucleotides of codon 76 and is 148 bp long. The third intron is located between codon 125 and 126 and is 304 bp long. All exon-intron junctions contain the expected consensus splice signal

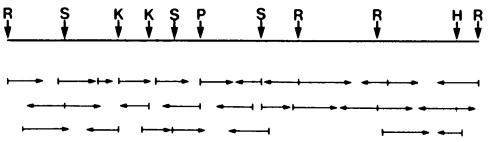

Figure 1. Sequencing strategem. Horizontal arrows show the direction and extent of sequences. Restriction sites at which DNA segments were subcloned into M13 mp8 or mp9 are indicated by single letters. H, Hind III; K, Kpn-I; P, Pst; R, Eco RI; S, Sma I.

dinucleotides GT (upstream) and AG (downstream).

The 3' end of the gene contains a TAA termination codon at position 452 and an AATAAA polyadenylation signal 192 bp further downstream. The characteristics of the 5' end of bα1 are discussed below.

The first intervening sequence of bα1 contains an arrangement of (GT) dinucleotide repeats. Two hundred and sixty five base pairs into the intron there are 18 (GT) dinucleotides flanked by a 26 bp direct repeat which is perfect except for a single base mismatch. An additional stretch of 12 (GT) pairs plus the sequence CTGTCTGTCTC is adjacent to the downstream 26 bp repeat and the entire arrangement of two poly(GT) sequences and the 26 bp repeat is itself encompassed by a 10 bp direct repeat that is also perfect except for a single base mismatch (Fig. 7A).

bα1 is Closely Related to a Rat α-Tubulin Gene

Several observations can be made on the basis of a structural comparison between the human bα1 gene and a rat α-tubulin gene to which it is related by extensive 3'-untranslated region homolgy (15,16). The number and position of the introns are identical in the rat and human genes, while the relative size of the introns is similar. The coding region sequences of the human and rat genes have been stringently conserved, and are 95°/∘ homologous at the DNA level. The 5°/∘ nucleotide sequence divergence occurs primarily in codon third positions and does not affect the encoded proteins which are absolutely identical in amino acid sequence. The 80°/∘ homology in the 3'-untranslated regions of the rat α-tubulin gene designated α-T14 and the human gene bα1 is depicted in a homology matrix plot in Fig. 3A.

Further comparison of the sequences of the rat and human α-tubulin genes revealed additional unexpected homologies. These homologies were found in the first intervening sequences of the two genes and are represented in a homology matrix plot in Fig. 3B. The homology occurs in a series of segments interrupted by non-homologous regions resulting in the broken diagonal seen

in the matrix plot. One block of homology (A) occurs between the beginning of IVS 1 in the human gene and a region 370 bp downstream from the beginning of IVS 1 in the rat gene. The sequences then diverge through a region

```
        GAATTCATGCCGTTGGGTGGAGTCAGCGCCCCCAGGCTCTACTTGGAAAACCTTTAAGCTCTTTTCTTTCGTAAGC
TCTCTGGGCGAGGGTGGTGGTATGTTTTGTGAGGTTTAGCTTAGCCCCAAATCCTCAAGCCCCGCCGCCGCCGCTAGTGCGGT
GCAGGAACCGGGCCAGTACTGCGCCCAGGGACAGAGCGCTGGGGAGGAACAAAGGCGGCGCTAGGCTGTGTTATCCGAGAGAT
CTTTCGGGGGCCGCGGGCAGCCCGTCCTGCCGCGACCGAGGGTCTGGGCGTCCCGGCTGGGCCCCGTGTCTGTGCGCACCGTT
TCGCTGATGCTGAGGGGCCACTTTCTGTCTCGCGTTGTTCTCTGGGGACCGGGAGAGGAGGAGGCACCCAAAAAGAGCGGGGG
CGTTGGGCGAGCTCGGGGGACGTGGGAGGGGGAACGGGAACAAAGCGCAGCCTAGGGTTAGCGTGGGAAGACCCTCCGCGGTC

                                                1
                                               Met
TTTGGCGTTTTGGAAAGATACCCACACATTCCCGGGAAAAC ATG GTGAGTTTCTGCCCGGAGCCCCCGGAGCGGGTGTCAG

GGCGGCGAGGGGCGGGGTTGTTTGTTTCTGGCTTCTATGGCGTTGGAGCCACTGGGCGCGGTTCGCCTCACTGAACCTCTTCT
GTCAGGAGCTGACTGAAAAAAAAACAAAAAAACCTTTCATCATTGCGGAACTGTAGGCTCCAAAAGGGTTTTCTTCACTATTA
TAAGTTAGATGACTTTTTTTTTTCTTGAGCAAAATCATAATTCACTTCACAAGCTCTTTAATGTCTGGTCTGGGGACGCCCTG
CCCTGACCGACTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGTGGGACGCCTGCCCTGACCGACTGAAGTG
TGTGTGTGTGTGTGTGTGTCTGTCTGTCTCGTCTGGACTGCACAGTTCAGCGAGGGAGAAAGGCCCACTTTGTGAGGGTAC
CGATGGTCAGGACCCAGGGAAACGCCCTTCCCCGCCGCCCCCCCGCCCCGCCCCCACCACATTCAGCGAATAGACAATTGAAA
GTGGTAGCCCTAAAGACCACAGAGAAGAAAACCTCTATTGGATGCAAAGAATATGAATATTATGTGATGGGTAGAGAATCTCA
GGATGAAAATACTATTTTGTTGTTTTAAATAAATATTTCATTATCCTTCCACTGGGCTTTTATTCTTTGGTACCTTTTCATGT
GATGCTTGTTTCTAACTTAGGAACTTTTGTGTGTGTGTGTGTAAGATACGGATAATTTTTCAGCTTTTACAGTGGAGAAGATC
TGGAAAAAGGTTTTTTTTTAAAAAAAAAAAGTAATGAAATTGCTACAGACAAAGAAGAATTATACTCCGCTTCCCGTTGTCCCC
CGTTCCAGTGCATCTTAATTAATTCATTTCAATTCAGGCACATGGTCCCGGGCGGTCAGAGGAGGAAAACTGGCAAAACAGCA
CAATGAGATCATGTAGGCAGCTGCTGGAAATAGAGCTTGCTCTGTTAAATAATGTAGCAGACAGTACAGGCTAGCACCAGGCA
CACAGCAAATACAGCAATGCAGCAATGCAGAAGGCAGACCTTGTCTAAACTCCTAGTATTGATGGATTCTGCAGTACAGATGT
CCGGATTATAATATCAAGTCCTATTCAGAGGAAACTTTCATCTTTATTTAAAAGGGAAGAAAGCAGTAAAATTAATCCCAATT
AAGTCATAATTGGATTTACTTCATTTTACAAATTTGTGCTTTGAATACTGAATAGCTTTTAAATATGAAAATCTTCTATTCAA
GACTGGTAGTAGGCCAATGGCTGGTACCCGTGCTGACAGGGCCAAGGCGACAATCATTATTCAGACCACACCCATATGCAGCA
TTTGTAGCAGGTGATTTTCCTTAAATCTTTGTATCGTGCTGGGGATATGACCTCAAATAATTTAGAAAAATATCTGTATATTA

                                                                                      2
                                                                                     Arg
TTAGAAATATTTTGAAATTTCCTATAATTTAAATGCTAATACACCTTAATTTTACATTTTTCACTTTTCCTCCCCACAG CGT

                                    10                                          20
Glu Cys Ile Ser Ile His Val Gly Gln Ala Gly Val Gln Ile Gly Asn Ala Cys Trp Glu Leu
GAG TGC ATC TCC ATC CAC GTT GGC CAG GCT GGT GTC CAG ATT GGC AAT.GCC TGC TGG GAG CTC

                                30                                          40
Tyr Cys Leu Glu His Gly Ile Gln Pro Asp Gly Gln Met Pro Ser Asp Lys Thr Ile Gly Gly
TAC TGC CTG GAA CAC GGC ATC CAG CCC GAT GGC CAG ATG CCA AGT GAC AAG ACC ATT GGG GGA

                            50                                          60
Gly Asp Asp Ser Phe Asn Thr Phe Phe Ser Glu Thr Gly Ala Gly Lys His Val Pro Arg Ala
GGA GAT GAT TCC TTC AAC ACC TTC TTC AGT GAG ACG GGG GCT GGC AAG CAT GTG CCC CGG GCA

                        70
Val Phe Val Asp Leu Glu Pro Thr Val Ile
GTG TTT GTA GAC TTG GAA CCC ACA GTC ATT G GTGAGTTGACCTCAGTAACCCAAGTGAGATCCCAGGGTGCT

GGACAGGAGGTCTGTCCTGGGGGGCTCCGCTGGTCACTCACCCACTCTCCTCCCGCTCCTCGTCCCCTCCTCCTCCTCCCCCT

                                                    80                              90
                                Asp Glu Val Arg Thr Gly Thr Tyr Arg Gln Leu Phe His Pro Glu
GCTCCTCCCCCATCATGTCTCCAG AT GAA GTT CGC ACT GGC ACC TAC CGC CAG CTC TTC CAC CCT GAG

                                            100                                     110
Gln Leu Ile Thr Gly Lys Glu Asp Ala Ala Asn Asn Tyr Ala Arg Gly His Tyr Thr Ile Gly
CAA CTC ATC ACA GGC AAA GAA GAT GCT GCC AAT AAC TAT GCC CGA GGG CAC TAC ACC ATT GGC

                                    120
Lys Glu Ile Ile Asp Leu Val Leu Asp Arg Ile Arg Lys Leu
AAG GAG ATC ATT GAC CTC GTG TTG GAC CGA ATT CGC AAG CTG GTATGTTTCTTTTCAAGAATAAAGTAA

ATTAATGAGCCTAAAGAACACTTTTGAAATAATGCTTTTTTTTTCAAACACAGAATTGAACTGTTATTTTAATAAAGAGTGGA
ATGAGTCATTCTTTGGGGTTTTTAAAATTCAGTTAAAATGAACTATTTGATGTCATTTTGTAAATGTTAATGAGAATTTTTTA
AAAGCATTTGTCAAATAAGATCTAAGTCCTGGAGATGTATGAAAGTGAAATATATTACTATGATGTACTACAAGATAAACTAA
```

```
                                                130
                                Ala Asp Gln Cys Thr Gly Leu Gln Gly Phe Leu Val Phe His
CCTTTCCTCTGTCCTCTCTTTTGTATAG GCC GAC CAG TGC ACG GGT CTC CAG GGC TTC TTG GTT TTC CAC

140                                         150                                 160
Ser Phe Gly Gly Gly Thr Gly Ser Gly Phe Thr Ser Leu Leu Met Glu Arg Leu Ser Val Asp
AGC TTT GGT GGG GGA ACT GGT TCT GGG TTC ACC TCG CTG CTC ATG GAA CGT CTC TCA GTT GAT

                                        170                                 180
Tyr Gly Lys Lys Ser Lys Leu Glu Phe Ser Ile Tyr Pro Ala Pro Gln Val Ser Thr Ala Val
TAT GGC AAG AAG TCC AAG CTG GAG TTC TCT ATT TAC CCG GCG CCC CAG GTT TCC ACA GCT GTA

                                    190                                 200
Val Glu Pro Tyr Asn Ser Ile Leu Thr Thr His Thr Thr Leu Glu His Ser Asp Cys Ala Phe
GTT GAG CCC TAC AAC TCC ATC CTC ACC ACC CAC ACC ACC CTG GAG CAC TCT GAT TGT GCC TTC

                                210                                 220
Met Val Asp Asn Glu Ala Ile Tyr Asp Ile Cys Arg Arg Asn Leu Asp Ile Glu Arg Pro Thr
ATG GTA GAC AAT GAG GCC ATC TAT GAC ATC TGT CGT AGA AAC CTC GAT ATT GAG CGT CCA ACC

                            230                                 240
Tyr Thr Asn Leu Asn Arg Leu Ile Gly Gln Ile Val Ser Ser Ile Thr Ala Ser Leu Arg Phe
TAT ACT AAC CTG AAT AGG TTA ATA GGT CAA ATT GTG TCC TCC ATC ACT GCT TCC CTG AGA TTT

                        250                                 260
Asp Gly Ala Leu Asn Val Asp Leu Thr Glu Phe Gln Thr Asn Leu Val Pro Tyr Pro Arg Ile
GAT GGA GCC CTG AAT GTT GAC CTG ACA GAA TTC CAG ACC AAC CTG GTG CCC TAT CCC CGC ATC

                    270                                 280
His Phe Pro Leu Ala Thr Tyr Ala Pro Val Ile Ser Ala Glu Lys Ala Tyr His Glu Gln Leu
CAC TTC CCT CTG GCC ACA TAT GCC CCT GTC ATC TCT GCT GAG AAA GCC TAC CAT GAA CAG CTT

                290                                 300
Ser Val Ala Glu Ile Thr Asn Ala Cys Phe Glu Pro Ala Asn Gln Met Val Lys Cys Asp Pro
TCT GTA GCA GAG ATC ACC AAT GCT TGC TTT GAG CCA GCC AAC CAG ATG GTG AAA TGT GAC CCT

            310                                 320
Arg His Gly Lys Tyr Met Ala Cys Cys Leu Leu Tyr Arg Gly Asp Val Val Pro Lys Asp Val
CGC CAT GGT AAA TAC ATG GCT TGC TGC CTG TTG TAC CGT GGT GAC GTG GTT CCC AAA GAT GTC

        330                                 340
Asn Ala Ala Ile Ala Thr Ile Lys Thr Lys Arg Thr Ile Gln Phe Val Asp Trp Cys Pro Thr
AAT GCT GCC ATT GCC ACC ATC AAG ACC AAG CGT ACC ATC CAG TTT GTG GAT TGG TGC CCC ACT

350                                         360                                 370
Gly Phe Lys Val Gly Ile Asn Tyr Gln Pro Pro Thr Val Val Pro Gly Gly Asp Leu Ala Lys
GGC TTC AAG GTT GGC ATC AAC TAC CAG CCT CCC ACT GTG GTG CCT GGT GGA GAC CTG GCC AAG

                                        380                                 390
Val Gln Arg Ala Val Cys Met Leu Ser Asn Thr Thr Ala Ile Ala Glu Ala Trp Ala Arg Leu
GTA CAG AGA GCT GTG TGC ATG CTG AGC AAC ACC ACA GCC ATT GCT GAG GCC TGG GCT CGC CTG

                                    400                                 410
Asp His Lys Phe Asp Leu Met Tyr Ala Lys Arg Ala Phe Val His Trp Tyr Val Gly Glu Gly
GAC CAC AAG TTT GAC CTG ATG TAT GCC AAA CGT GCC TTT GTT CAC TGG TAC GTT GGG GAG GGG

                                420                                 430
Met Glu Glu Gly Glu Phe Ser Glu Ala Arg Glu Asp Met Ala Ala Leu Glu Lys Asp Tyr Glu
ATG GAG GAA GGT GAG TTT TCA GAG GCC CGT GAG GAC ATG GCT GCC CTT GAG AAG GAT TAT GAG

                            440                                 450
Glu Val Gly Val Asp Ser Val Glu Gly Glu Gly Glu Glu Glu Gly Glu Glu Tyr
GAG GTT GGT GTG GAT TCT GTT GAA GGA GAG GGT GAG GAA GAA GGA GAG GAA TAC TAAAGTTAAAA

CGTCACAAAGGTGCTGCTTTTACAGGGAAGCTTATTCTGTTTTAAACATTGAAAATGTTGTGGTCTGATCAGTTAATTTGTAT
GTAGCAGTGTATGCTCTCATATCAATTACTGACCTATGCTCTAAAACATGAATGCCTTTGTTACAGACCCAAGCTGTCCATTT
CTGTGATGGGTTTTGAATAAAGTATTCCCTGTCTTAAATGAATTC
```

Figure 2. Sequence of the bαl gene derived by the strategem shown in Figure 1. The cap site and poly(A) addition signal are boxed; poly(GT) elements and direct repeats in IVS 1 are dashed; three short sequences in the 5' flanking region with homology to sequences in the 5' end of a rat gene (α-T14, ref. 15) are underlined.

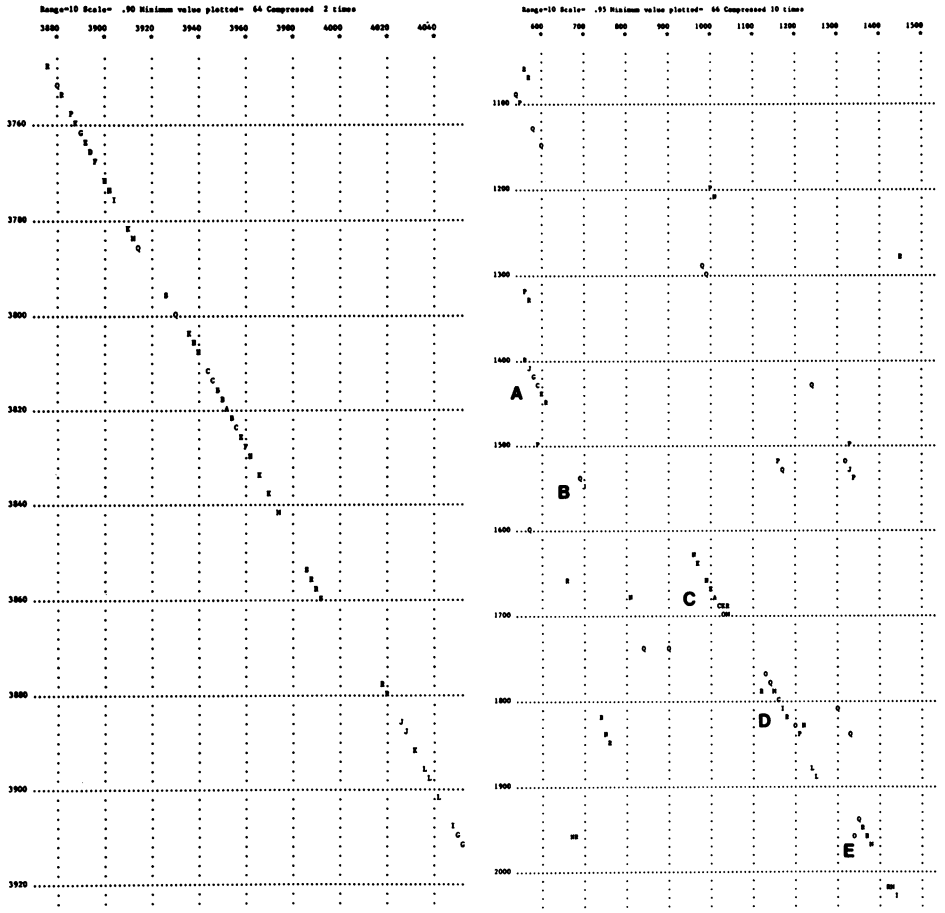

Figure 3. Homology matrix plots generated by the Pustell forward homology matrix computer program (22). The abscissa corresponds to the human bαl sequence; the ordinate corresponds to the rat α-T14 sequence. Numbers on the axes represent the positions of nucleotides in each sequence. Homologies between the sequences over a range of 10 bp are indicated by a letter. Each letter signifies a percent homology (e.g., A = 100°/°, F = 90-91°/°, K = 80-81°/°). In order to plot homologies across regions of DNA that contain more bases than the number of columns on a page, the sequences are compressed. The compression factor refers to the number of bases assigned to each coordinate in the plot. Figure A depicts homology between the 3' untranlsated regions of bαl and α-T14; the minimum percent homology plotted is 64°/°. The compression factor is 2. Figure B depicts homology between the first intervening sequence of bαl and α-T14. Blocks of homology are designated A,B,C,D, and E. The minimum value plotted is 66°/°. The compression factor is 10.

Figure 4. S_1 protection of a 5'-end labeled restriction fragment by total IMR 6 poly(A)$^+$ mRNA. After annealing and S_1 digestion (see Methods), the reaction products were analyzed on 5 % polyacrylamide-urea sequencing gel. Lanes 1-4 represent the dideoxy sequence (T,C,G,A) of a clone used as a size marker. Lanes 5,6; no mRNA controls. Lanes 7 and 8; IMR 6 mRNA. 2000 units of S_1 nuclease were used for the reactions analyzed in lanes 5 and 7. 3000 units of enzyme were used for the reactions analyzed in lanes 6 and 8.

containing a A_9CA_7 tract in the human, but no corresponding oligo A tract in the rat. Homology resumes briefly (B) and then is lost again in the region of the human intron containing the (GT) dinucleotide repeats. The (GT) repeats are not present in the section of the rat intron that has been sequenced. Intermittent homology (C,D,E) continues 900 bp into the human and rat introns. Whether the homology persists throughout the remainder of the introns could not be established because the complete sequence of the rat intron was not available.

It should be noted that the segments of homology in the introns of the rat and human genes are composed of relatively complex sequences. Regions of the introns that contain simple sequences or repeats are not homologous. In contrast to the first introns, comparison of the second and third introns of the rat and human genes revealed no significant homology.

Characteristics of the 5' end of bαl

To determine the location of the 5'end of the bαl gene an S_1 nuclease protection experiment was performed (Fig. 4). A 5' end labeled restriction fragment spanning a 250 bp region upstream from the putative ATG was annealed

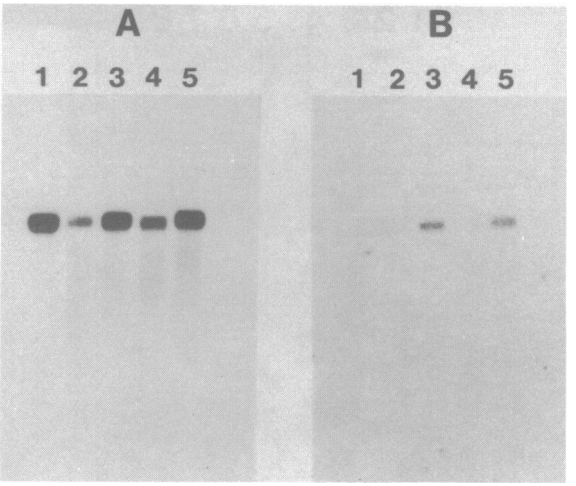

Figure 5. Restricted expression of bα1 to cell lines of neurological origin. Approximately 5 μg of poly(A)$^+$ mRNA from HeLa cells (Lane 1), Y79 retinoblastoma cells (Lane 2), 132 1N1 glioma cells (Lane 3), CHP 126 neuroblastoma cells (Lane 4), and IMR 6 neuroblastoma cells (Lane 5) were resolved on an agarose gel containing formaldehyde (23). The contents of the gel were transferred to nitrocellulose (24) and the blots were probed with nick-translated α-tubulin coding region (Figure A) or the bα1 3' untranslated region (Figure B). A summary of this and other bα1 expression data (16) appears in Table 1.

to polyA$^+$ RNA from a human neuroblastoma cell line that expresses the bα1 gene (see below). S_1 digestion yielded a cluster of protected fragments approximately 201 bp in length. The 5' end label of the protected fragment is 8 bp from the ATG, so the 5'-untranslated region is approximately 209 bp long. This result places the cap site of bα1 transcripts at the position indicated in Fig. 2.

The bα1 5' flanking sequence is unusual in that it does not contain a TATA box -35 bp upstream from the cap site. Surprisingly, the rat gene α-T14 5' end does contain a TATA box. Comparison of the 5' ends of bα1 and α-T14 reveals that the cap site of bα1 is approximately twice the distance from the initiation codon (209 bp) as the cap site of α-T14 (99 bp) and that, whereas the 5' flanking sequences are not closely related, there are distinctive homologies. These homologies are limited to three regions (underlined in Fig. 2 and shown in Fig. 7B). The regions are 13, 20, and 29 bp long and are respectively 92°/∘, 80°/∘, and 76°/∘ homologous. Region I is 341 bp upstream from the cap site in the rat gene and 279 bp upstream from the cap

Table I
Expression of bα1 in Various Human Cell Lines
and in Human Fetal Brain

HeLa cells	-	Retinoblastoma Y79	-
Diploid fibroblasts	-	Glioma 132 1N1*	+
Epidermal cells	-	Neuroblastoma CHP 126	-
Squamous cell carcinoma SCC-15	-	Neuroblastoma IMR 6*	+
Fibrosarcoma	-	Neuroblastoma IMR 32*	++
Brain (fetal)	+++	Neuroblastoma CHP 134d*	++
Myeloma CM 1500A	-	Neuroblastoma SKN-SH*	+
Hepatoma PLC/PRF 15	-		

Cell lines marked with an asterisk are adherent.

site in the human gene. Region II is 286 bp upstream from the cap site in the rat gene and 261 bp upstream from the cap site in the human gene. The third region of homology is 121 bp upstream from the cap site in the rat gene; in the human gene this sequence is transcribed.

Restricted Expression of the bα1 Gene

In previous experiments, the bα1 3'-untranslated region subclone was used as a probe for RNA blot transfer experiments using polyA$^+$ RNA from the following human cell types: HeLa cells, diploid fibroblasts, epidermal cells, a squamous cell carcinoma, a fibrosarcoma, and fetal brain. The only sample in which bα1 RNA was detected was human fetal brain (16). In order to extend these experiments, we have examined a variety of cultured human cell lines of neurological origin for bα1 expression. The resulting RNA blot transfer experiments are shown in Fig. 5. In the first panel, polyA$^+$ RNA from HeLa cells, a retinoblastoma (Y79), a glioma (1321N1), and two neuroblastomas (CHP 126 and IMR 6) were probed with an α-tubulin coding region sequence. In the second panel the same RNAs were probed with the bα1 3'-untranslated region subclone. α-tubulin mRNAs are present in all the cell lines. In contrast, bα1-specific mRNA is undetectable in HeLa cells, the retinoblastoma, and one of the neuroblastomas (CHP 126), but present in the glioma and the second neuroblastoma (IMR 6). The complete bα1 expression data, including results from three additional neuroblastomas are summarized in Table I.

DISCUSSION

Conserved Structure of α-Tubulin Genes

Southern blots of human DNA probed with the bα1 3'-untranslated region subclone show that the human genome contains two copies of this sequence both of which are associated with α-tubulin specific sequences. Two distinct but

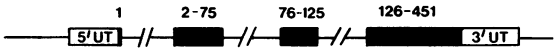

Figure 6. Block diagram of the exon-intron arrangement of mammalian α-tubulin genes known to be expressed. Numbers represent the amino acids encoded in each exon.

unrelated sets of overlapping recombinant clones containing each of these copies have been isolated. Sequence analysis of one of these clones revealed that within the α-tubulin coding region, homology to the bα1 cDNA is less than perfect and that point mutations in the sequence preclude the synthesis of a functional α-tubulin protein. Accordingly, this clone must be classified as a pseudogene. In the absence of complete sequence data, it is not clear whether this pseudogene contains intervening sequences, or is of the processed type. The only other genomic copy of the 3'-untranslated region is associated with α-tubulin coding region sequences that are completely homologous to the bα1 cDNA sequence. Hence, this copy belongs to a unique functionally expressed human α-tubulin gene.

The presence of α-tubulins with related functional and biochemical properties in all eukaryotic organisms is consistent with the evolutionary conservation of tubulin coding regions and protein sequences. The amino acid homology between the human bα1 and rat α-T14 tubulins is absolute. The human and rat sequences differ at only two residues from a pig α-tubulin sequence (26) and at three residues from a chicken α-tubulin sequence (27). The similarity between the rat and human genes extends to their structure as well and includes the conservation of both the number and placement of introns within the genes. This structural identity raises the question whether all vertebrate α-tubulin genes conform to the exon arrangement shared by bα1 and α-T14 (Fig. 6). Conservation of structure appears to be a feature of the β-tubulin genes of higher eukaryotes; all of the expressed human and chicken β-tubulin genes thus far characterized contain the same number of introns in the same locations (6). However, heteroduplex analysis of another human α-tubulin gene suggests that the carboxyterminal exon is split by an additional intervening sequence and therefore this gene represents an alternative to the structure shared by bα1 and α-T14 (28).

An interesting feature of the rat and human α-tubulin genes is the occurrence of an intervening sequence immediately after the first codon. This feature is not unique to α-tubulin genes. A single mouse α-amylase gene contains two 5' leader sequences both of which are spliced after the

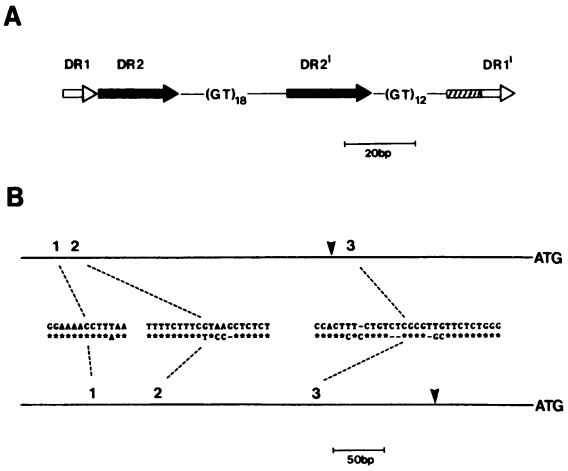

Figure 7. A. Block diagram (drawn to scale) showing arrangement of oligo(GT) sequences and direct repeats in the first intervening sequence of bα1. Hatched region represents a unique 9 bp sequence downstream from the $(GT)_{12}$. B. Location of three upstream homologous regions in bα1 (human; top) and α-T14 (rat, bottom) showing position relative to the cap sites (arrows) and initiation codons. Asterisks denote homology between the human and rat sequences.

initiator triplet (29). In this case the position of the first intron is associated with the flexible expression of the α-amylase gene. The two different leaders are employed in different tissues; one or the other leader is spliced onto the common coding region of the gene. There is no evidence for such a mechanism among α-tubulin genes, but the presence of an intervening sequence after the initiator ATG has the effect of isolating 5' regulatory sequences from the coding region of the gene and may provide the basis for other as yet undefined functions.

Interspecies Conservation of Intron Sequences

The substantial nucleotide sequence homology between the human bα1 and rat α-T14 genes extends to non-coding regions. The 3'-untranslated regions of these genes are extensively homologous and there is segmental homology in the first intervening sequences. Interspecies homology of intron sequences has been reported among globin genes and proto-oncogenes. For example, the introns of the goat and human ε-globin genes are 64-69°/∘ homologous (30); and the introns of the mouse c-fos proto-oncogene are 63-76°/∘ identical to the introns of the human cellular homologue c-fos (31). In light of the fact that intron sequences are not generally conserved even among members of a gene family within a species, these instances of interspecies homology of

intron sequences define a category of genes that are extensively conserved. It is not yet known how many genes belong in this category and it remains to be determined whether the homologies in IVS 1 of bα1 and α-T14 or the homologies in their 3'-untranslated regions are functionally significant.

The Large Intron of bα1 Contains a Repetitive Element

A notable feature of the bα1 gene is the presence of the oligo (GT) repetitive element in the first intron. Members of this family of dispersed repetitive sequences have been found in the genomes of yeast, Xenopus, mouse, and human (32-34). In the latter species, there are estimated to be 30,000-50,000 copies per genome. The example cited here is interesting in that it contains two (GT) stretches and two sets of flanking direct repeats (Fig. 7A). The direct repeats associated with these (GT) dinucleotides are sufficiently complex to qualify as duplicated sequences rather than as simple sequences that happen to surround the oligo(GT) elements. Evidently, poly(GT) elements occur in two genomic environments: some are flanked by direct repeats whereas others are not (34,35). The presence of the flanking direct repeats suggests that these sequences are mobile genetic elements, and is consistent with the insertion of the (GT) tracts at sites generated by staggered chromosomal breaks.

Proposals regarding the function of these repeats are numerous (34): they may function at the ends of chromosomes, they may serve as recombination hotspots, they may be involved in gene conversion events, or they may influence the transcription of genes with which they are associated. The latter view includes the notion that poly(GT) sequences might have enhancer properties by virtue of their ability to form left-handed DNA helices or Z-DNA (36).

Evolution of the bα1 Promoter

Current theories of gene regulation have stressed the importance of relatively short sequences located upstream from the cap site of regulated genes (37). These sequences, which are distinct from the TATA box and the CAAT "boxes", have been demonstrated to be essential for the transcriptional activation of a variety of eukaryotic genes. In some cases, the regulatory sequences (which may be as short as 9 bp) are repeated several times in the 5' flanking region of a gene. Their identification has been facilitated among genes that are coordinately expressed. For example, the family of genes that are induced by glucocorticoids all share a 23 bp upstream consensus sequence (38). A 16 bp consensus sequence is repeated in the 5' flanking region of Chlamydomonas reinhardi α- and β-tubulin genes which are

also transcribed coordinately (39). Short regulatory sequences have also been identified upstream from genes that are not known to be coordinately induced (40).

From this perspective, the three short regions of homology in the 5' ends of bα1 and α-T14 (Fig. 7B) are candidates for regulatory sequences in α-tubulin gene expression. A segment of region 3 is particularly interesting. The sequence of this element is TTCTCTGGG. In both the rat and human it is followed by an AG-rich sequence. Interestingly, this element followed by an AG-rich sequence also occurs in the 5' end of the human β-tubulin gene designated 5β (5). This element is, therefore, a potential regulatory sequence for both α- and β-tubulin gene expression and could conceivably provide a basis for coordinate regulation of α- and β-tubulin gene expression. A demonstration of the influence of these short sequences on tubulin gene expression will require the construction of appropriate mutants, and the introduction of such constructs into host cells by DNA-mediated gene transfer.

Considering the overall similarity of the human bα1 and rat α-T14 genes, it is striking that bα1 does not contain a TATA box whereas α-T14 does. Comparison of these two genes suggests that bα1, an otherwise remarkably conserved gene, has diverged in its promoter region. Pressure for a divergent promoter could not have come from a requirment for differential regulation that might be associated with the emergence of a new α-tubulin isotype, since the bα1 and α-T14 proteins are identical. Rather, there is the possibility that evolutionary pressure was applied to the 5' regulatory elements themselves. Accordingly, the divergent promoter of bα1 may be an example of the sort of evolution recently suggested for the tubulin gene families by Raff: separate regulatory mechanisms for functionally equivalent genes might evolve because of different requirements for tubulin transcripts in different cells or at different developmental stages (41).

bα1 Expression

The data presented here suggest that the expression of bα1 is regulated at the level of tissue specificity. Expression of bα1 has only been detected in cells of neurological origin. However, since our expression data are based on RNA blot transfer analysis we cannot rule out the possibility of extremely low levels of bα1 expression in other cell types. In contrast, transcripts of the rat gene α-T14 are prominent in fibroblasts as well as brain cells (15).

The variable expression of bα1 among the different cell lines of

neurological origin suggests a correlation between bα1 expression and cellular morphology. Cells in which bα1 mRNA levels are undetectable, such as the retinoblastoma Y79 and the neuroblastoma CHP 126, are small round cells that grow in suspension. Cells in which the bα1 gene is expressed are adherent neuroblastoma or glioma cells with characteristic cytoplasmic processes. The exclusive identification of bα1 transcripts in morphologically differentiated cells raises the question whether the differentiation of these cells involves the regulated expression of the bα1 gene. In order to address this issue, we are assessing transcription of the bα1 gene in retinoblastoma and neuroblastoma cells in which differentiation has been induced in tissue culture.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Health (GM 31740) and the Muscular Dystrophy Association.

REFERENCES

1. Kirschner, M.W. (1978) Int. Rev. Cytol. 54, 1-71.
2. Cleveland, D.W., Lopata, M.A., McDonald, R.J., Cowan, N.J., Rutter, W.J. and Kirschner, M.W. (1980) Cell 20, 95-105.
3. Wilde, C.D., Crowther, C.E., Cripe, T.P., Gwo-Shu Lee, M. and Cowan, N.J. (1982) Nature 292, 83-84.
4. Gwo-Shu Lee, M., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell 33, 477-487.
5. Gwo-Shu Lee, M., Loomis, C. and Cowan, N.J. (1984) Nucl. Acids Res. 12, 5823-5836.
6. Lewis, S.A., Gilmartin, M.E., Hall, J.L. and Cowan, N.J. (1984) Submitted to J. Mol. Biol.
7. Ben Ze'ev, A., Farmer, S.R. and Penman, S. (1979) Cell 17, 319-325.
8. Cleveland, D.W., Lopata, M.A., Sherline, R. and Kirschner, M.W. (1981) Cell 25, 537-546.
9. Bond, J.F., Robinson, G.S. and Farmer, S.R. (1984) Mol. Cell. Biol. 4, 1313-1319.
10. Lopata, M.A., Havercroft, J.C., Chow, L.T. and Cleveland, D.W. (1983) Cell 32, 713-724.
11. Kemphues, K.J., Raff, R.A., Kaufman, T.C. and Raff, E.C. (1979) Proc. Natl. Acad. Sci. USA 76, 3993-3995.
12. Kemphues, K.J., Raff, E.C., Raff, R.A. and Kaufman, T.C. (1980) Cell 21, 445-451.
13. Distel, R.J., Kleene, K.C. and Hecht, N.B. (1984) Science 224, 68-70.
14. Havercroft, J.C. and Cleveland, D.W. (1984) Submitted to J. Cell Biol.
15. Lemischka, I.R. and Sharp, P.A. (1982) Nature 300, 330-335.
16. Cowan, N.J., Dobner, P.R., Fuchs, E.V. and Cleveland, D.W. (1983) Mol. Cell. Biol. 3, 1738-1745.
17. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1980) Cell 19, 959-972.
18. Ozaki, L.S., Kimura, A., Shimada, K. and Takagi, T. (1981) J. Biochem. 91, 1155-1162.
19. Sanger, F., Coulsen, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B. (1980) J. Mol. Biol. 143, 161-178.

20. Hu, N. and Messing, J. (1982) Gene 17, 271-277.
21. Staden R. (1980) Nucl. Acids Res. 8, 3673-3694.
22. Pustell, J. and Kafatos, F.C. (1984) Nucl. Acids. Res. 12,
23. Boedtker, H. (1971) Biochim. Biophys. Acta 240. 448-453.
24. Southern, E. (1975) J. Mol. Biol. 98, 503-517.
25. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
26. Ponstingl, H., Little, M., Krauhs, E. and Kempf, T. (1981) Proc. Natl. Acad. Sci. USA 78, 2757-2761.
27. Valenzuela, P., Quiroga, M., Zalduar, J., Rutter, W.J., Kirschner, M.W. and Cleveland, D.W. (1982) Nature 289, 650-655.
28. Wilde, C.W., Chow, L.T., Wefald, F.C. and Cowan, N.J. (1982) Proc. Natl. Acad. Sci. USA 79, 96-100.
29. Young, R.A., Haenbuchle, O. and Svhibler, U. (1981) Cell 23, 451-458.
30. Shapiro, S.G., Schon, E.A., Townes, T.M. and Lingul, J.B. (1983) J. Mol. Biol. 169, 31-52.
31. Straaten, F.V., Muller, R., Curran, T., Beveren, C.V. and Verma, I.M. (1983) Proc. Natl. Acad. Sci. USA 80, 3183-3187.
32. Miesfeld, R., Krystal, M. and Arnheim, N. (1981) Nucl. Acids Res. 9, 5931-5947.
33. Hamada, H. and Kakunaga, T. (1983) Nature 298, 396-399.
34. Rogers, J (1983) Nature 305, 101-102.
35. Sun, L., Paulson, K.E., Schmid, C.W., Kadyk, L. and Leinwand, L. (1984) Nucl. Acids Res. 12, 2668-2690.
36 Rich, A., Nordheim, A. and Wang, A. (1984) Ann. Rev. Biochem. 53, 791-846.
37. Davidson, E.H., Jacobs, H.T. and Britten, R.J. (1983) Nature 301, 468-470.
38. Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4867-4871.
39. Brunke, K.J., Anthony, J.G., Sternberg, E.J. and Weeks, D.P. (1984) Mol. Cell. Biol. 4, 1115-1124.
40. McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.
41. Raff, E.C. (1984) J. Cell Biol. 99, 1-10.